

THE EFFECTS OF STIMULUS MODALITY AND TASK INTEGRALITY:
PREDICTING DUAL-TASK PERFORMANCE AND WORKLOAD FROM SINGLE-TASK LEVELS

Sandra G. Hart Robert J. Shively Michael A. Vidulich
NASA-Ames Research Center
Moffett Field, CA

Ronald C. Miller
Informatics General Corporation
Palo Alto, CA

ABSTRACT

The influence of stimulus modality and task difficulty on workload and performance was investigated in the current study. The goal was to quantify the "cost" (in terms of response time and experienced workload) incurred when essentially serial task components shared common elements (e.g., the response to one initiated the other) which could be accomplished in parallel. The experimental tasks were based on the "Fittsberg" paradigm; the solution to a SternBERG-type memory task determines which of two identical FITTS targets are acquired. Previous research suggested that such functionally integrated "dual" tasks are performed with substantially less workload and faster response times than would be predicted by summing single-task components when both are presented in the same stimulus modality (visual). In the current study, the physical integration of task elements was varied (although their functional relationship remained the same) to determine whether dual-task facilitation would persist if task components were presented in different sensory modalities. Again, it was found that the cost of performing the two-stage task was considerably less than the sum of component single-task levels when both were presented visually. Less facilitation was found when task elements were presented in different sensory modalities. These results suggest the importance of distinguishing between concurrent tasks that compete for limited resources from those that beneficially share common resources when selecting the stimulus modalities for information displays.

INTRODUCTION

The current experiment is one in a series that investigated the rules by which single task estimates of workload or performance can be used to predict the results of different task combinations. Theoretically, some task combinations should be simply additive; the workload of two tasks performed concurrently should be equal to the sum of component task levels. This was found, for example, by Gopher and Braune (1984). In this study, as in many others, however, performance on one or both of the component tasks suffered when they were presented concurrently. Numerous experiments have been conducted with a dual-task paradigm in which a variety of tasks are presented and learned individually and then different combinations are performed concurrently. It is assumed that subjects' resources can be allocated, up to their limit, in graded quantities among separate activities. The fact that some tasks appear to interfere with each other more than others led to the formulation of a multiple resources model that postulated that different amounts and types of resources are required for different tasks and task combinations (Navon & Gopher, 1979). Performance limitations arise from insufficient resources in one or more processes that might be differentiated by the

modality of input, output, or type of central processing (Wickens & Kessel, 1979). In many cases, the difficulty levels of one or both tasks are varied to determine the limits of capacity (Kantowitz & Knight, 1978). In addition, the required performance levels or task emphasis may be specified (Gopher, Brickner and Navon, 1982) to shift the relative priorities among dual-task components. It was found that subjects can dynamically allocate their attention to achieve the required levels of performance (Tsang & Wickens, 1984).

The dual-task paradigm has been used to identify the causes and magnitudes of dual-task performance decrements and subjective workload experiences with different combinations of input and output modalities, levels of loading, and requirements for stages of cognitive processing. In general, it has been found that performance on one (or both) tasks suffers to the extent the demands for resources exceeds the system capacity (Wickens, Sandry and Vidulich, 1983). For example, the decrement in performance for a visual/manual spatial transformation task was found to be greater than for the same task presented with auditory input and speech output when each was performed with a visually displayed manual control task (Vidulich & Tsang, 1985a; 1985b). This occurred even though the auditory/manual version of the spatial transformation task was performed more slowly and imposed more workload when presented as a single task. Subjective workload ratings for the dual-task combinations were somewhat less than the sum of the single-task levels. However, the cost (in terms of subjective workload experience) was significantly greater for dual-task combinations with the same input and/or output modalities, than for those that were presented in different sensory modalities or required responses in different output modalities. Dual-task workload ratings were equal to 60% of the sum of single task levels for tasks with different input or output modalities, and 75% of the sum of single-task levels for tasks that competed for the same resources.

The results of dual-task experiments, particularly those within the general structure of multiple resources theory, have provided ideas and guidance for design engineers faced with the problem of off-loading visually (or manually, vocally, etc) overloaded operators with alternative information sources or response modalities. For example, voice input or synthesized voice output has become an almost universal proposal for off-loading pilots whose ability to process additional visual information has been exceeded (Vidulich and Wickens, 1985). In addition, graphic display alternatives have been proposed to replace digital displays of instruments and the need for information integration has been recognized in order to reduce the physical number of sources and formats of information (National Research Council, 1983). Not all concurrent task components can be divided among different sensory modalities with the same improvements in performance and workload, however. It is possible that task elements that are functionally related by the structure of the task or their temporal relationship should be presented or performed in the same input or output modalities, while unrelated but concurrent tasks should be displayed or performed in different sensory modalities. The former might promote subjective integration, thereby reducing workload (Wickens & Yeh, 1982; 1983), whereas the latter can reduce competition for limited resources, also reducing workload.

In the typical dual-task paradigm, the two tasks must be performed within the same time period (thereby competing for an operator's limited resources), yet the component tasks are unrelated either functionally or subjectively. An alternative paradigm would be one in which component tasks are functionally related; the output or response to one serves to initiate or provide information for the other. This type of task is common in operational environments where the decision to initiate a change in a system's state requires preliminary information gathering, processing, and decision making, which is followed by one or more discrete or continuous control actions. The sources of information, processing requirements, response modality, and workload levels of the first stage are independent of those of the second stage. Nevertheless, the two tasks are functionally related and some or many processing stages may either be performed in parallel, or the activities required for one may simultaneously satisfy

some of the requirements of the other. For example, mental anticipation and physical response preparation for a control input can begin while instruments are monitored to determine the correct value or time for the control input. For these types of tasks, it is possible that presenting information in the same sensory modality would result in reduced workload and dual-task performance time, which is in direct opposition to the typical dual-task finding.

The tasks selected for the current study were based on the "Fittsberg" paradigm (Hartzell, Gopher, Hart, Dunbar, & Lee, 1983) in which a target acquisition task based on FITTS Law (Fitts & Petersen, 1964) was combined with a SternBERG memory search task (Sternberg, 1969). Two identical targets are displayed equi-distant from a centered probe stimulus. Subjects acquire the target on the right if the probe is a member of the memory set and the target on the left if it is not. Performance on the response selection portion of the task is evaluated by measures of speed (reaction time - RT) and accuracy (percent correct and decision reversals). Response execution is accomplished by moving the control stick in the selected direction (right or left) and acquiring the target on the selected side of the display. Target acquisition performance is evaluated by measuring movement time (MT), which is the total time required to acquire the target less RT. Target acquisition difficulty is manipulated within blocks of trials by varying the width (W) of the target area and its distance from the home position of the cursor (A) according to Fitts' Law ($MT = a + b(ID)$) where:

$$\text{Index of Difficulty (ID)} = \log_2(2A/W)$$

MT, but not RT, increased as the difficulty of the target acquisition task was increased. RT but not MT increased as the cognitive load of the response selection task was increased. Subjects rated the workload of the combined "Fittsberg" task as slightly greater than the workload of the response selection task by itself. Workload ratings for a block of trials in which different levels of target acquisition difficulty were imposed integrated the load levels imposed by both the response selection and response execution components.

In subsequent experiments (Hart, Sellers & Guthart, 1984; Mosier & Hart, 1985; Staveland, Hart & Yeh, 1985), response selection was accomplished by responding to directional commands presented symbolically or with linguistic abbreviations, identifying a stimulus with or without the additional task of comparing it to a remembered value, computing the results of mathematical equations, performing matching tasks, and time estimation, among others. The response selection demands ranged from none (in the single-target Fitts baseline condition) to stimulus identification, short-term or long-term memory search, prediction, computation, comparison, and estimation. Again, the two-stage "Fittsberg" tasks were performed with approximately the same performance and rated workload as the response selection tasks performed alone. A small "concurrency cost" (Navon and Gopher, 1979) of 40 msec in RT was again found for the combined tasks, as well as a slight increase in rated workload over single task levels (from 33 to 43). Dual task RTs were equal to 63% of the sum of single task levels and dual task workload ratings were equal to 64% of the sum of single task levels. MT was never affected by response selection difficulty manipulations. Again in opposition to the results of traditional dual-task experiments, performance decrements for the response selection (measured by RT) or response execution components (measured by MT) were not found as the difficulty of the other component was increased. Rather, the two components appeared to impose independent (or at least parallel) demands that did not increasingly degrade performance as load levels of one or both was increased.

Although this could be considered a dual-task paradigm, the response selection and execution elements can be performed sequentially and their difficulty manipulated independently, in

keeping with the assumptions of serial models of memory scanning (Sternberg, 1969) and information theoretic models of choice reaction time and target acquisition (Fitts & Petersen, 1964). In addition, the types of activities that are represented are typical of many operational environments in which operators must decide what to do (response selection) and then accomplish the desired function (response execution). The results of earlier studies suggest that the addition of automation to accomplish one or more functions might have limitations in effectiveness to moderate the demands placed on busy operators. If the execution of control inputs is automated, this might simply reduce the response execution load, leaving the demands of response selection (e.g., when and how to initiate the system) unchanged and providing little real savings in performance time or workload for functionally integrated tasks.

The current experiment was designed to address one of the issues raised earlier: For functionally integrated tasks, is the savings (measured in terms of workload, response time, or accuracy) found for functionally related tasks presented in the same sensory modality also present when the same tasks are presented in different sensory modalities? Four response-selection tasks were presented individually (in the single-task baseline experiment) and in combination with a target acquisition task (in the dual-task, Fittsberg experiment): (1) right/left decision based on spatial (Spatial); (2) or linguistic (Right/Left) information; (3) Sternberg memory search with a memory set size of one (Memory-1); and (4) Sternberg memory search with a memory set size of four (Memory-4). Each response selection task was presented visually and auditorially in both baseline and Fittsberg experiments. In the Fittsberg experiment, each response selection task was coupled with visually displayed target acquisition tasks.

The goal was to determine the rules by which dual-task performance and workload levels might be predicted from single-task levels. The spatial and linguistic command conditions were included to determine whether the large RTs found for a Right/Left condition in two earlier studies (Hart et al, 1984; Hartzell, et al, 1983) occurred because a directional command presented with a verbal code (R or L) was more difficult to translate into a directional movement than a spatial command or because additional time was required to translate the abbreviation (R or L) into its linguistic representation (right or left). The two levels of memory task difficulty were included to investigate the possibility of an interaction for measures of performance and workload between stimulus modality and the subsequent processing requirements for probes that were identical in meaning but not physical representation.

The specific experimental predictions were:

1. For simple right/left decision tasks, spatial stimuli will result in faster RTs and lower workload ratings, replicating earlier studies.
2. For memory search tasks, RT and workload will be directly related to memory set size, replicating earlier studies.
3. MT will be unaffected by the difficulty or modality of the response selection task, replicating earlier studies.
4. For both single- and dual-task presentations, the auditory display modality will result in slower RTs and higher workload ratings
5. When response selection and response execution task components are presented in the same sensory modality, substantially more dual-task facilitation will be found than when they are presented in different modalities.

METHOD (Single-task and Dual-Task Experiments)

Subjects

Eight subjects, five men and three women participated in the single-task baseline study. None of them had served in earlier Fittsberg experiments. Eight different subjects, six men and two women served as paid participants in the dual-task experiment. All of them had served previously in an experiment in which they had received extensive training on the target acquisition task coupled with many different response selection tasks.

Apparatus

The experiment was conducted in a small experimental booth. Subjects were seated in a chair located 85 cm from a 23-cm monitor where the experimental tasks were displayed. The visual angle subtended by the most extreme targets was 11 degrees. A two-axis joystick was mounted on the right arm of the chair for the response selection and target acquisition responses. Workload-related rating values were selected with a slide-pot and entered with a button mounted on the left arm of the chair. The experiment was performed with an Apple II+ microcomputer and a Cyborg ISAAC interface modified to allow rapid and accurate recording of responses (to the nearest 10 msec). Subjects wore stereo headsets to receive stimulus information for the auditory response selection conditions. Tones were generated by the ISAAC. Linguistic information for the Right/Left and Memory tasks was generated by a Votrax Type n' Talk.

Experimental conditions

The basic task involved a binary decision to move to the right or left. The stimulus for the visual response selection tasks was a single symbol (< or >), alphabet letter (e.g., "A", "D", etc), or word ("Right" or "Left") presented in the center of the display. Stimuli for the auditory response selection tasks were presented via stereo headphones. Tones for the spatial task were presented monaurally to either the right or left ears. Right/Left commands, the memory set item(s), and memory task probes were presented binaurally. For the Fittsberg experiment, two identical targets were symmetrically presented on either side of the screen at the onset of the response selection task. (Figure 1) Their distance from the center (A) was determined by the ID for that trial. (Figure 1) The targets were two 1.25 cm vertical lines separated by the distance (W) specified by the ID for the trial. A 1.25 cm vertical line (the cursor) was controlled by movement of the joystick.

Response selection Tasks

The baseline experiment provided single-task performance and workload comparisons for the dual-task experiment. Each response selection task was presented as a choice reaction time task in both auditory and visual modalities. There were four levels of response selection difficulty: (1) Spatial command; (2) Right/Left command; (3) Memory-1; and (4) Memory-4. For the dual-task experiment, the cursor and targets were presented visually at the same time that either auditory or visual response selection stimuli were initiated.

A/Spatial information was generated by the ISAAC system. A short tone burst (1000 Hz) was presented for 1000 msec in either the right or left ear cuff. V/Spatial information was presented immediately beneath the centered cursor: "<" and ">" for left and right movement respectively. A/Right/Left commands were generated by a Votrax Type n'Talk speech synthesizer. The word "Right" or "Left" was presented binaurally at the beginning of each trial. Utterance durations were 400 and 500 msec respectively. For V/Right-Left trials, the word "Right" or

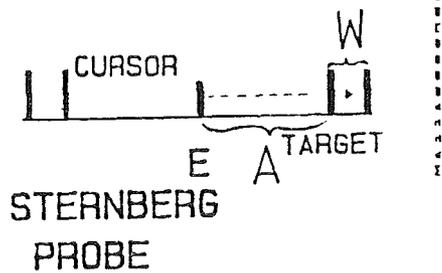


Fig. 1 Typical Fittsberg display
Memory set = 4

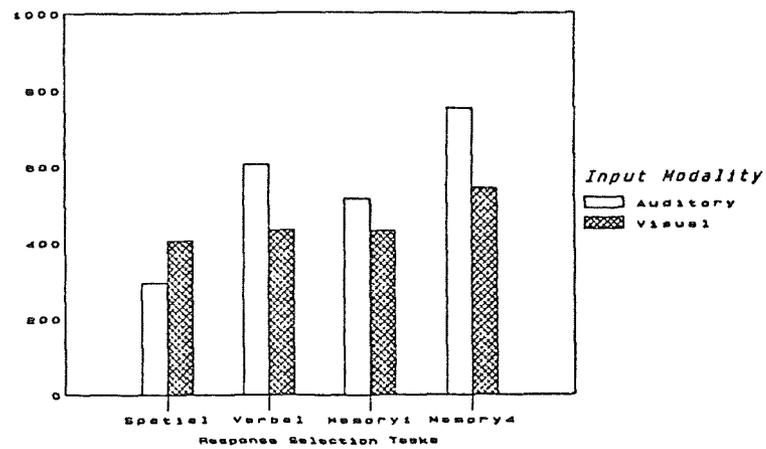


Fig. 2 Reaction times - Single Task Conditions

"Left" was displayed centered beneath the cursor. A/Memory trial blocks were preceded by binaural presentation of the memory set item(s) for the entire block of trials (e.g., "A" might be presented for Memory-1; and "B", "M", "T" and "R" for Memory-4) generated by the Votrax. Single-letter probes, also generated by the Votrax, were presented at the onset of each trial. The average duration of the alphabet-character stimuli was 300 msec. For V/Memory trials, letters were displayed on the CRT for 2000 msec before each block of trials and centered beneath the cursor at the beginning of individual trials. In the visual modality response selection stimuli remain on the display until the trial is completed.

Response execution

Response Selection component. The interval between onset of the response selection stimulus and a 2% stick deflection to the right or left was recorded as the RT. RT intervals were computed from stimulus onset for both auditory and visual presentations, as the total time required to process information is the most operationally relevant measure to use in comparing alternative stimulus presentation modalities.

Target acquisition component. The combinations of target widths and amplitudes used were all that were possible within the limited precision of the display (widths ranged from 5 to 20 pixels, amplitudes from 60 to 128 pixels). Three IDs were created (2.52 (40/60), 4.19 (either 7/64 or 14/128), and 5.67 (5/128)) in accordance with Fitts' Law. They were the same IDs that were used in earlier experiments. They were randomly presented within each block of 24 experimental trials (mean ID = 4.15). MTs were recorded as the interval between the end of the response initiation portion of the task (RT) until the steadiness criterion for keeping the cursor within the selected target had been satisfied. Single-task baseline levels for the target acquisition tasks were obtained by randomly presenting one of the four possible target configurations on the right or left.

Knowledge of results

Immediately after each trial ended (either by the selection of a response or by target acquisition), the experimental display was replaced for 2 sec by a verbal evaluation of RT and MT performance (if the subject had made a correct decision) or the word "WRONG" (if the subject selected an incorrect direction of movement). The verbal evaluations (e.g., "Fantastic", "Good", "Truly Dismal", etc.) were based on norms obtained in earlier studies.

Rating Scales

Workload experiences were evaluated by computing a derived score (Hart, et al, 1984) based

on evaluations of nine workload-related factors obtained after each experimental condition, weighted to reflect the importance placed on the factor by individual subjects. The nine factors were considered to be representative of the dimensions considered relevant to different individuals' definitions of workload: task difficulty (TD), time pressure (TP), own performance (OP), physical effort (PE), mental effort (ME), frustration (FR), stress (ST), fatigue (FA), and activity type (AT).

The relative importance of the nine factors to each subject (e.g. the weights) was determined by a pretest. All possible pairs of the nine factors were presented on the computer display in a different random order to each subject. The member of each pair selected as most relevant to workload was recorded and the number of times each factor was selected was computed. The resulting values could range from 0 (not relevant) to 8 (more important than any other factor).

Subjects rated their experiences after each experimental condition on the same nine workload-related dimensions and a single global rating of workload. Each scale was presented on the experimental display as a 11-cm vertical line with a title (e.g. "MENTAL EFFORT") and bipolar descriptors at each end (e.g. "EXTREMELY HIGH/EXTREMELY LOW"). Numerical values from 0 to 100 were assigned to the selected scales positions during data analysis.

Procedure

A brief introduction was read to familiarize subjects with the purpose of the study and the types of tasks they were to perform. Then, the workload weights were obtained. The eight experimental conditions were presented in a counter-balanced order to the subjects in both experiments. Each condition consisted of 72 trials; two blocks of 24 practice trials presented immediately before a block of 24 experimental trials. For all conditions, half of the correct responses were "right" and half were "left", and were presented in random order. The bipolar rating scales were presented after completion the third block of experimental trials. The baseline study required one, two-hour session. The Fittsberg experiment required two three-hour sessions.

RESULTS AND DISCUSSION

Single-Task Baseline Experiment

The following data were obtained: percent correct, average RT, and bipolar ratings for each block of experimental trials. Individual 2-way and 1-way analyses of variances for repeated measures were performed between experimental conditions to determine if the predicted changes in performance and workload occurred due to response selection difficulty and stimulus modality. Selected correlations were performed among the raw bipolar ratings, weighted workload scores, and RT.

Percent Correct

Responses were made relatively accurately; average values ranged from 84% to 98% across subjects and from 87% (V/Memory-4) to 98% (V/Spatial) across experimental conditions. The difference in accuracy for the four response selection tasks was statistically significant ($F(3,21) = 6.18, p < .01$). Although slightly more correct responses were made for the auditory display modality, the difference was not significant ($F(1,7) = 3.99, p > .10$). These differences were in the same direction as the reaction times, thus ruling out the possibility of a speed-accuracy trade-off (Pachella, 1974).

Reaction Time

There were highly significant differences in RT among the response selection tasks ($F(3,21) = 50.44, p < .001$) and stimulus modalities ($F(3,21) = 45.74, p < .001$). (Figure 2) However, there was a significant interaction between the two variables ($F(1,7) = 28.10, p < .001$). RTs were 170 msec faster for the spatial tasks than for any other conditions. For this task, RT was considerably faster for the auditory mode of presentation than for the visual mode. A tone presented in one ear or the other is an imperative stimulus having immediate directional connotations that apparently required a minimal level of processing for a directional decision to be completed. For the Right-Left and Memory tasks, however, RTs were as much as 200 msec faster for the visual mode of presentation than for the auditory mode. The same difference occurred in RT between spatial and linguistic presentation of a directional command that was found in the earlier studies, suggesting that the earlier results were not due to difficulty in translating an abbreviation (e.g., R for right) into the word it represented. Rather, the increase in RT reflected difficulty in translating a linguistic command into a spatial movement.

It is unlikely that the presentation time for auditory stimuli influenced the modality differences. Not only was the RT shorter for the Spatial condition, but the magnitude of the differences for the remaining conditions was great enough, that the effect could not be explained by stimuli durations, although a potential confound exists. RT was recorded from the onset of the stimulus presentation. Thus, while the visual information was immediately available, the temporal nature of the auditory stimuli does not allow immediate information extraction. However, identification of information does not require the entire stimulus interval to be completed Remington (1977).

Relative importance of workload-related factors (Weights)

Subjects' initial biases about the factors they would consider in evaluating workload were obtained in a pre-test. Figure 3) Even though there was considerable diversity among the subjects' opinions, as expected, there was a small but statistically significant difference in the average importance placed on the nine factors ($F(8,56) = 3.41, p < .01$). Mental Effort was the most important factor, while Physical Effort and Fatigue were the least. There was the most disagreement about the importance of Frustration and Activity Type. A multiple correlation was performed on the weights. The only statistically significant positive correlations found were for Stress (with Time Pressure and Fatigue). The only significant negative correlations found were for Activity Type (with Frustration, Time Pressure, and Stress). These results suggest that, not only do subjects disagree about the relative importance of different factors to workload, but there are few consistent relationships among the factors themselves.

Workload Ratings/Derived Score

Bipolar ratings obtained after the third replication of each experimental condition varied widely in average values and standard deviations across subjects and experimental conditions: TD (24/16); TP (40/24); OP (41/24); ME (32/16); PE (8/11); FR (38/23); ST (35/20); FA (32/22); AT (15/18); and OW (26/15). Not only did subjects disagree about what factors were relevant to workload, but they also disagreed about the degree to which each of the factors were imposed by or experienced during different experimental conditions (e.g., standard deviations were occasionally greater than the average values).

Following the procedure used in earlier studies, (Hart et al, 1984; Vidulich & Tsang, 1985) a derived workload score was computed that reflected the subjective importance of each factor for each subject. Factors that were essential to an individual's concept of workload might be entered many times whereas others, considered less important, might be entered few times, or not at all. The averaged combination of the weighted ratings was used as the primary measure of subjective workload. As has been found in every other application of this technique,

significant relationships among experimental variables (estimated by previous research, Overall Workload ratings, and performance measures) were maintained or increased, while average between-subject variability within each experimental condition was decreased. In this experiment, between-subject standard deviations, within experimental conditions, were reduced from 14 to 11.

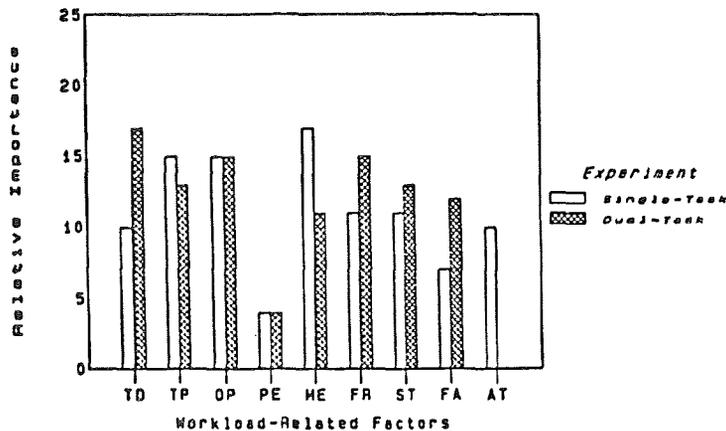


Fig. 3 Relative importance of workload related tasks Single vs. Dual Tasks

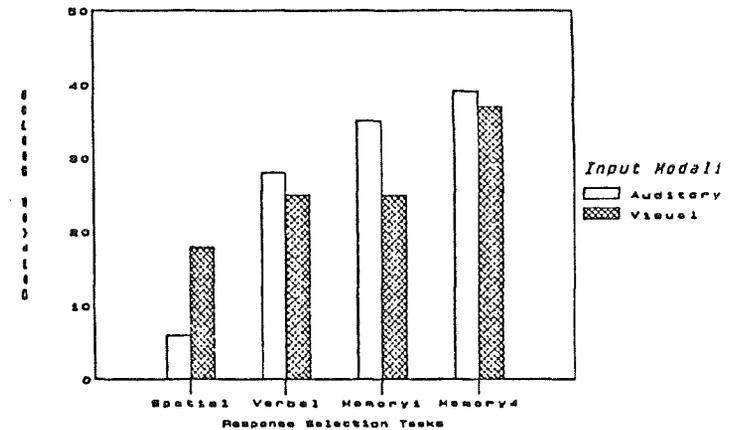


Fig. 4 Weighted Bipolar Workload Ratings Single-Task Conditions

The Derived Workload Score reflected a pattern of statistical significance similar to that obtained for RT. There was a significant difference among the four experimental tasks ($F(3,21) = 15.52, p < .001$), and a significant interaction between display modality and response selection task ($F(3,21) = 3.19, p < .05$). The spatial decision task was less loading in the auditory modality whereas the visual versions of the other tasks were more loading. (Figure 4) As expected, the spatial decision task was considered less loading than the Right/Left decision task ($F(1,7) = 9.65, p < .05$) and a memory set size of one for the Sternberg task was experienced as less loading ($F(1,7) = 5.51, p < .05$) than a memory set size of four.

The relationships among the individual scales, and their association with overall workload (the weighted workload score) and RT were determined by a multiple correlation. The nine workload-related factors were not independent, suggesting a potential source of problem for multi-dimensional rating scale techniques that require statistical independence among the dimensions. Task Difficulty and Stress were related to many other factors whereas Activity Type was not. Task Difficulty, Own Performance, Mental Effort, Frustration, and Stress were significantly correlated with Overall Workload ratings and all of the factors were significantly correlated with the Derived Workload score. Although the latter result may be an artifact of the weighting procedure, it possibly reflects the fact that the derived score represents a composite of factors relevant to each subject, providing a common denominator across subjects (regardless of the factors that each considered) and measuring the workload imposed by a specific task. Few rating scales were significantly correlated with RT, even though both measures were significantly influenced by experimental manipulations. In fact, Task Difficulty and Overall Workload were the only scales that even approached a significant relationship. This finding again points out the importance of obtaining independent measures of workload and performance, as they may reflect different phenomenon.

Dual-Task, Fittsberg Experiment

The following data were analyzed: percent correct, number of decision reversals, average

RT and MT, and bipolar ratings for each block of experimental trials. Preliminary one-way analyses of variance for repeated measures were performed within blocks of trials to examine differences in performance attributable to the direction of response. Two-way analyses of variance for repeated measures were performed between experimental conditions to determine whether the predicted changes in performance and workload occurred, and multiple correlations were performed to assess the associations among bipolar ratings, derived workload scores, and performance measures.

Direction of Movement

There were no significant differences in correct selections or MT between targets presented on the right or left. There was a significant right/left differences in RT for the memory tasks (but not the other response selection tasks), as expected; "yes" responses (to the right) were made significantly more quickly ($F(1,7) = 8.02, p < .05$) than "no" responses (to the left). This is a common finding with the Sternberg paradigm (Sternberg, 1969). Since there was an equal number of right and left conditions and because it did not interact with any of the other experimental variables, subsequent analyses were performed without regard for the direction of movement.

Percent Correct

The number of incorrect response selections did not vary significantly across experimental conditions ($F < 1.0$) or stimulus modalities ($F < 1.0$). Since errors were made on less than 2% of the trials, there appears to be no evidence of a speed/accuracy tradeoff. Somewhat more reversed decisions were found. A reversed decision is one in which initial decision (identified by the direction of movement recorded for RT) was made in a different direction than the target that is acquired subsequently. The differences were statistically significant for memory set size ($F(1,7) = 10.66, p < .01$) and spatial versus linguistic directional command ($F(1,7) = 17.14, p < .01$). Spatial commands resulted in 2.5 times fewer control reversals (less than 1 per block of 24 trials) than linguistic commands (2.5 per block). Finally, a significant interaction was found between Stimulus Modality and Method of Presentation for the direction command tasks ($F(1,7) = 7.00, p > .05$). There were more reversals for V/Right-Left than A/Right-Left (4 versus 2 per block) whereas both A/Spatial and V/Spatial conditions were performed with consistently few reversals (less than 1 per block), regardless of stimulus modality. Subsequent analyses for performance measures included non-reversed trials only, to eliminate very long MTs for trials in which reversed decisions occurred.

Reaction Time

RTs for the dual-task conditions were generally lower ($F(1,14) = 20.75, p < .05$) reflecting differences in abilities between the two groups of subjects. However, there was no interaction between experiment and response selection manipulations.

RT differences within the dual-task experiment were similar to those obtained for the baseline experiment, providing sensitive indicators of response selection manipulations. (Figure 5) There was a highly significant difference in RT among the four response selection tasks ($F(3,21) = 34.83, p < .001$). The expected differences were found between the spatial and linguistic presentation modes for the direction tasks (345 msec vs 442 msec) and between the difficulty levels of the memory task (422 msec vs 528 msec). In addition, there was a significant difference between stimulus modalities: responses to visual stimuli were generally made more quickly than to auditory stimuli ($F(1,7) = 11.62, p < .05$). There was, however, a significant interaction between stimulus modality and response selection task ($F(3,21) = 43.73, p < .001$), as was found in the Baseline experiment. RTs were slower for the V/Spatial than for the A/Spatial tasks, whereas the other tasks were performed more quickly with visual information than auditory.

RT for the target acquisition task presented in its single-task configuration was 421 msec, virtually the same time required to perform the simplest response selection/target acquisition task presented in the dual-task mode (413 msec), and within 100 msec of the most difficult task (Memory-4). Since the response selection tasks required at least 296 msec (A/Spatial) and as much as 754 msec (A/Memory-4) to complete by themselves, it is clear that some of the processing required to complete the response selection portion of the Fittsberg task and the initial preparation for target acquisition must have progressed in parallel. In every case, the obtained performance was equal to one half or less of the levels that would be predicted by simply adding the single task levels. This finding replicates that of earlier studies.

To adjust the reaction time distributions for the two different population samples (Experiment 1 versus Experiment 2) the following transformation was performed. Each distribution was converted to z-scores based upon its own mean and standard deviation. A grand mean was then computed on both distributions and the variances were pooled. The original z-scores were then multiplied by the square root of the pooled variance and added to the grand mean. This produces a single distribution with a mean based on all data, while retaining the shapes of the original distributions. When this transformation was applied, significant overall differences were found for response selection and stimulus modalities (as found for the experiments individually), but no interaction was found between either of these factors and experiment. When RT for the dual-task was predicted with these transformed scores, obtained RTs were 49% of the sum of single task levels for the visual modality and 60% of the sum of single task levels for the auditory modality; a significant difference in the cost of performing complex but functionally related tasks.

Movement Time

Although MTs were not analyzed within each block of trials to determine whether or not the linear relationship predicted by Fitts Law between ID and MT held, it was assumed that it did, as the same set of target configurations had been used in all of the earlier experiments, where this relationship was found. MTs for the three IDs were combined within each trial block for subsequent analyses, as each ID occurred the same number of times and no interaction between target ID and response selection difficulty manipulations was found in any of the earlier studies. No significant differences in MTs due to direction of movement were found for any of the experimental conditions.

Single-task baseline MTs averaged 888 msec. In contrast, average MTs for the Fittsberg, dual-task conditions, ranged from 834 to 874 msec across experimental conditions, 100 to 150 msec faster than were obtained in earlier studies and within 48 msec of the baseline level. (Figure 6) As predicted, there was no significant difference among MTs due to response selection load. There was, however, a significant difference in MT due to the modality of the response selection task ($F(1,7) = 11.41, p < .01$); MTs were significantly longer when the decision of which of two targets to acquire was presented auditorially than when it was presented in the same visual modality as the target acquisition task itself. These differences were observed for every response selection task, ranging from 10 to 100 msec. Thus, there was no interaction between response selection tasks and modality ($F < 1.0$). This is the first time that MT differences have been found due to response selection manipulations for any of the Fittsberg experiments. It is also the first time that the response selection tasks were presented auditorially as well as visually. It is possible that there is an extra cost (in MT) for processing and responding to information presented in one modality and then completing a subsequent task presented in another. This increase in MT following auditory presentation of a response selection task occurred even though the output for the response selection task (which initiated movement toward the correct target) was completed before the MT interval began. These results were based on correct and non-reversed decisions and, therefore, did not occur as a result of inaccuracy or indecision.

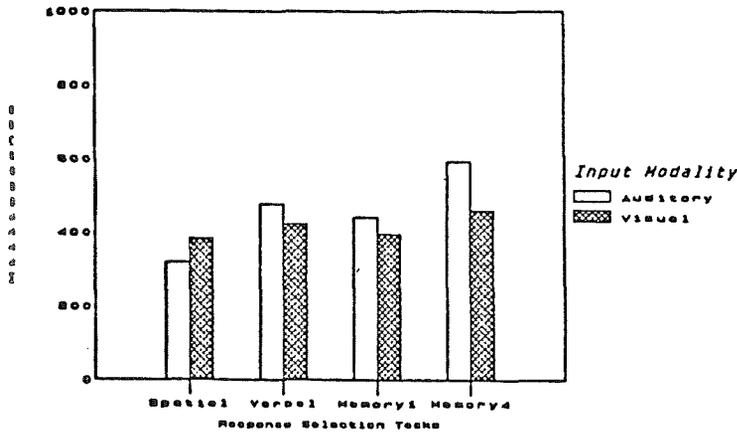


Fig. 5 Reaction Times
"Fittsberg Conditions"

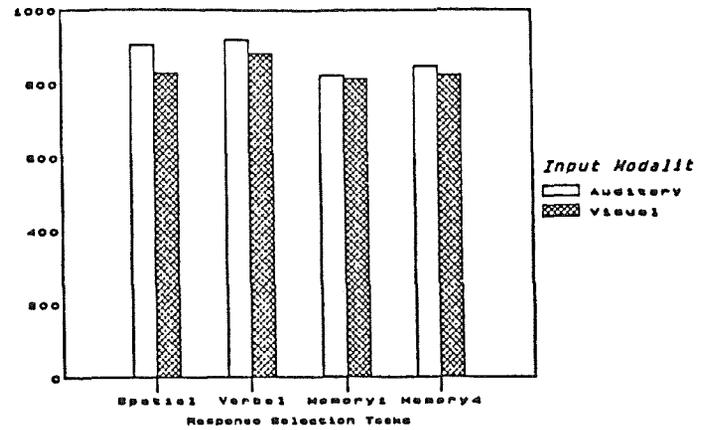


Fig. 6 Movement Times
"Fittsberg Conditions"

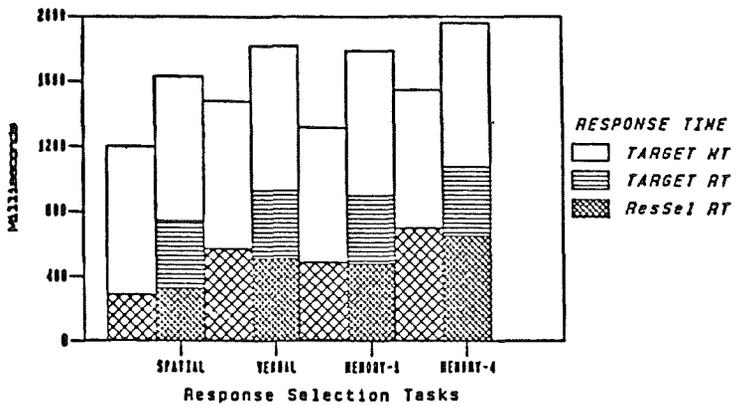


Fig. 7a Auditory/Visual

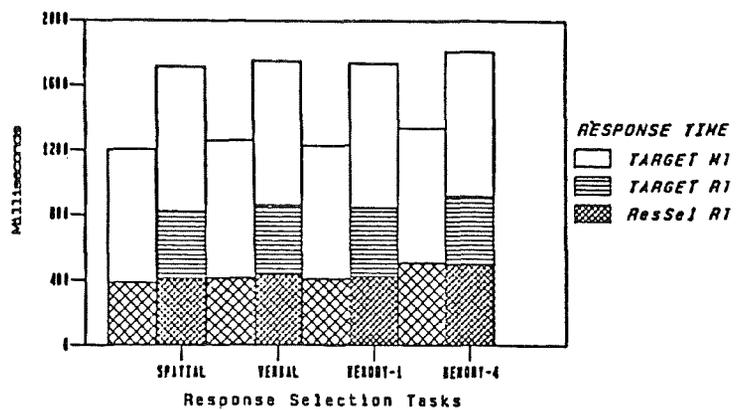


Fig. 7b Visual/Visual

$Total\ Time\ (RT(RS) + (RT(RE) + RE)$
 Leftmost Bars - Obtained Total Times
 Rightmost Bars - Predicted Times (Sum of Single Tasks Levels)

Because MTs were influenced by response selection modality, it is not clear whether all of the initial preparation required to perform a visual target acquisition (as estimated by RT in the single-target baseline condition) was completed in parallel with and by the end of a response selection decision if it was based on auditory information. Although target acquisition preparation could have been transferred to the beginning of the MT interval, given the design of the Fittsberg paradigm, this does not appear to have occurred in earlier studies, nor did it occur in the current study. Single-task baseline levels for MT were only 888 msec, 45 msec slower than the average dual-task MTs. Thus, this can not account for a significant portion of the 300-500 msec difference in predicted dual-task RTs compared to the sum of the single task levels and the obtained dual-task values.

Total Response Time

The total response time is the interval between stimulus presentation and target capture (the sum of RT and MT). Total times ranged from 1200 to 1440 msec across experimental conditions.

These values ranged from 70 to 81% of the levels that would be predicted by combining single-task target-acquisition RT, MT, and response selection RT for each condition. The predicted and obtained total times may be seen in Figure 7a,b. As you can see, there was a significant difference due to stimulus modality ($F(1,7) = 20.75, p < .001$) and response selection task ($F(3,21) = 12.89, p < .001$) when the two measures of performance were combined. There was no significant interaction. Obtained levels for the visual modality were 71% of the predicted levels and 77% for the auditory modality; again a reliable difference in the cost of performing complex but functionally integrated tasks presented in the same or different modalities.

Relative importance of workload-related factors (Weights)

The importance placed on eight of the workload-related factors may be seen in Figure 3. The Activity Type scale was not used, since it had demonstrated so little relationship with experimental manipulations in the earlier study. For this reason, only 28 pairwise combinations of factors were evaluated and the maximum value that any factor could assume was 7 (rather than 8). As you can see, there were large differences among subjects, although Task Difficulty, Own Performance, and Frustration were selected significantly more often than the rest ($F(7,49) = 3.04, p < .01$). There was the greatest agreement among subjects about Physical Effort and the least agreement about Time Pressure and Fatigue. Again, a correlation matrix was obtained to determine the relationships among the individual factors. No statistically significant correlations were found. The weights for the eight factors in common between the two experiments were compared to determine the degree of similarity between the two groups of subjects. The two groups were not found to be significantly different. They agreed that Physical Effort and Fatigue were relatively unimportant and that Frustration, Task Difficulty, Stress, and Own Performance were important. Although the differences were not statistically significant, the two groups disagreed about the importance of Frustration, Fatigue, and Mental Effort.

Workload Ratings/Derived Score

Again, there were large differences among subjects in the degree to which subjects that felt different factors were present in specific experimental conditions. The grand mean and overall standard deviations for the nine scales were: TD (24/17); TP (22/13); OP (29/17); ME (25/18); PE (10/12); FR (21/19); ST (20/18); FA (13/18); and OW (22/18).

Following the procedure used in the first experiment, bipolar ratings were weighted to compute a derived workload score. The weighted bipolar ratings were compared to those obtained in the baseline experiment. There was a highly significant difference ($F(1,14) = 26.63, p < .001$) between the magnitudes of ratings in the two experiments; they were consistently larger in the single-task experiment (33) than in the dual-task experiment (21), although between-subject standard deviations were identical. This may either reflect fundamental differences in the two groups of subjects, or a difference in the level of experience each had with the Fittsberg paradigm. The dual-task subjects had many hours of practice with the target acquisition tasks and a variety of response selection conditions. Thus, their perception of the workload imposed by the specific conditions included in this study could have been influenced by their previous experiences. Despite this difference, there were no significant interactions between experimental group and experimental manipulations ($F < 1.0$).

Workload ratings followed the same pattern obtained in the baseline experiment and for RTs. As you can see in Figure 8, there was a significant difference in experienced workload among the response selection tasks ($F(3,21) = 7.13, p < .01$). The most demanding task was the Memory-4 task (29). The least demanding task was the Spatial task (14). In addition, there was a significant difference due to stimulus modality ($F(1,7) = 13.18, p < .01$); auditory was generally rated as more loading (23) than visual (19). In addition, a significant

interaction between stimulus modality and response selection task was found ($F(3,21) = 13.34$, $p < .001$), in agreement with the first experiment and RT performance; the Spatial task was less loading when presented auditorially, whereas the other tasks were more loading. As expected, the spatial presentation of the directional task was significantly less loading than the Right/Left version ($F(1,7) = 9.52$, $p < .01$) and the Memory-1 was significantly less loading than Memory-4 ($F(1,7) = 5.29$, $p < .05$), replicating earlier results.

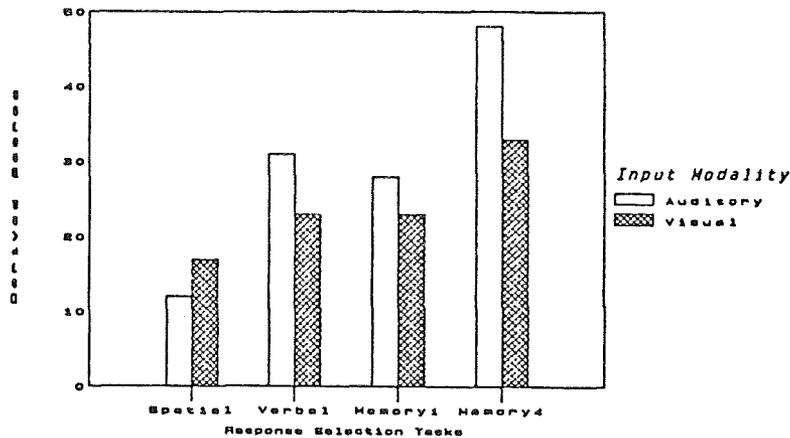


Fig. 8 Weighted Bipolar Workload Ratings
"Fittsberg Conditions"

The correlations among the nine bipolar ratings, weighted workload ratings, and total response time were obtained. Again, there was wide variation in the degree to which the different scales covaried with each other. Most of the individual scales were significantly correlated with each other with the exception of Own Performance, which was independent of the other scales. The dimensions were not, obviously, orthogonal. Every scale except Own Performance was significantly correlated with Overall Workload and all scales were significantly correlated with the derived workload scores, as was found before. None of the subjective measures were significantly correlated with total time, although they had each reflected many of the same experimental manipulations individually. This finding provides additional support for the suggestion that there may be a dissociation between measures of workload and performance (Wickens & Yeh, 1982; 1983).

Because the basic levels of ratings in the two experiments were so different, they were transformed employing the technique described earlier for RTs. When this transformation was applied, the ratings from the two experiments could be compared more directly. No significant interactions between experiment and experimental manipulations were found. Dual-task workload levels were equal to approximately half of the sum of single-task levels for the Spatial tasks (A and V). For the remaining, tasks, visual/visual conditions were equal to 49% of the baseline task sum while auditory/visual conditions were equal to 61% of the baseline task sum. This suggests that there was greater savings (in workload experienced) with tasks presented in the same sensory modality than for those presented in different modalities (Figure 9a,b).

CONCLUSIONS

This experiment succeeded in answering a number of questions about the influences of response selection and response execution difficulty and modality on measures of performance and workload. As has been found in earlier experiments with the Fittsberg paradigm, response selection load significantly affected RT but not MT. Both RT and MT were significantly longer

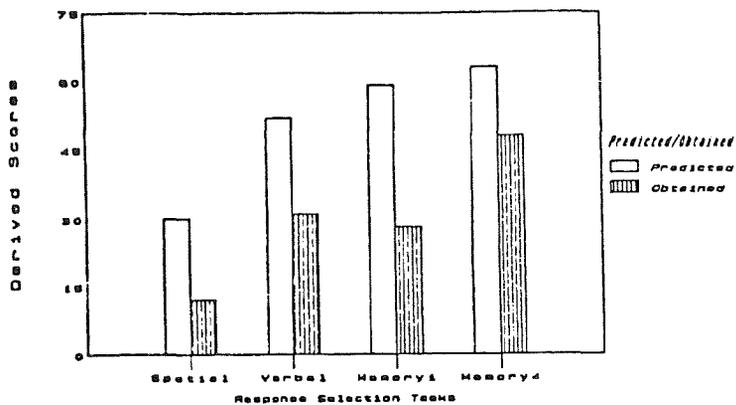


Fig. 9a Auditory/Visual

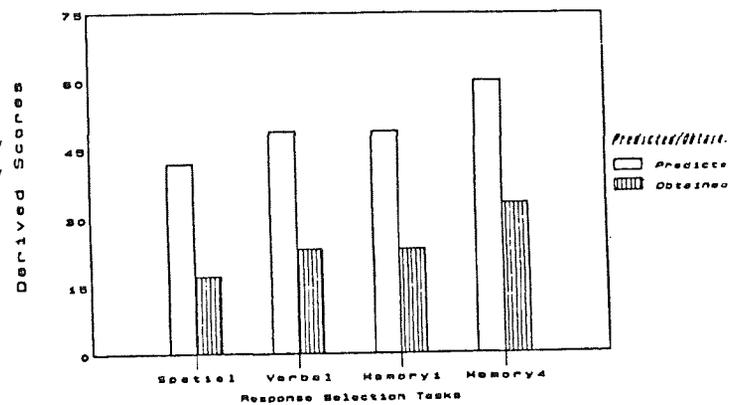


Fig. 9b Visual/Visual

Obtained vs Predicted Workload Ratings
Obtained = Dual-Task "Fittsberg" Ratings
Predicted = Sum of Single-Task Ratings

when linguistic information required for response selection was presented in a different sensory modality than the subsequent response execution task. The number of correct responses did not discriminate between any of the response selection tasks, however the frequency of reversed decisions did. The weighted averaged bipolar ratings were significantly influenced by both response selection and response execution difficulty manipulations and the stimulus-modality compatibility of the two task components.

Even though there were significant and consistent patterns of performance and workload changes as a result of all experimental manipulations, the correlations among the different measures were not statistically significant. This reinforces the point made by Wickens and Yeh (1982; 1983) that measures of workload and performance may dissociate as each is particularly sensitive to different, often subtle, aspects of experimental manipulations. For example, in the current study, both measures were sensitive to the modality of input and the response selection load, although there was an interaction between stimulus modality and difficulty for workload ratings but not for total response time or percent correct. These factors were independently influenced by each experimental manipulation. For this reason, subjective evaluations as well as multiple measures of performance are desirable to obtain a complete understanding of task demand characteristics.

Difficulty manipulations for one or both task components did not result in an interaction for any measures of performance or workload between single- and dual-task presentations. Such an interaction might have been expected with a traditional dual-task paradigm. This could have occurred because the capacity of the subjects was not exceeded by the task requirements (although there was a small RT and workload cost for putting the two tasks together), but this concurrence cost was consistent across difficulty manipulations and did not interact with level of difficulty. This provides additional support for the assertion that specific types of task combinations result in different patterns of performance and workload (e.g., either interference or facilitation).

Workload ratings integrated all task elements; both response selection and response execution sources of loading were both represented in subjective evaluations. In addition, ratings were sensitive to differences in the workload imposed by the alternative stimulus modalities, as were measures of speed and accuracy. This occurred even though there was considerable disagreement among the subjects about which dimensions were considered when evaluating workload and about the absolute magnitudes of these factors during any specific task.

As expected, the visually presented response selection tasks were well integrated with the visual target acquisition components. This physical stimulus compatibility enhanced the functional integration inherent in the Fitts design (e.g., the output for one served to initiate the other). The result was a considerable savings in response time and experienced workload over what might have been expected by combining single task load or duration levels. In general, RTs were 49% of the predicted additive levels, total times were 71% and workload ratings were only 46%. Response preparation for the Fitts target acquisition portion of the task was either performed in parallel with (or was replaced by) the response selection requirements of the combined tasks.

For the auditory display modality, however, the savings were not as great. RTs were 60% of predicted levels, total response times were 77%, and workload ratings were 56%. In addition, the requirement to switch from processing an auditory stimulus (in the response selection task) to acquiring a visually presented target imposed an additional cost of as much as 100 msec that was reflected in increased MTs. This could have occurred because of a modality switching cost. Alternatively, the fact that the visual stimuli remained on the display during target acquisition allowed reconfirmation of response selection during this phase, whereas auditory stimuli ended before target acquisition began, thereby requiring echoic memory for reconfirmation. Although all of these values were still less than the sum of single task levels, the savings in performance time and workload were not as great. For response selection tasks that shared the least processing requirements with the response execution task (e.g., the Memory-4 task), the obtained values approached 80% of the levels predicted by adding single task levels. For this task, the additional requirement of a four-item memory search task (particularly when conducted with auditory stimuli) required a significant amount of time and effort on the part of the subjects, yet only the final decision of "yes" or "no" was directly related to the subsequent target acquisition.

These results would not be predicted in traditional dual-task paradigms where it is commonly found that concurrent tasks presented in different sensory modalities impose less interference and workload, and those in the same modalities, more. Instead, it was found that both functional and physical integration of task components resulted in a facilitation of performance and a reduction in rated workload that were often less than either single-task level. These results suggest the importance of evaluating the relationships among task components when considering display modalities in operational environments. It would appear that concurrent but independent tasks would be best presented in different sensory modalities to reduce the competition for resources if stimulus/response compatibility is not grossly violated. For task elements that are functionally related, however, the opposite might be true. Task components should be presented in the same sensory modality to enhance an operator's ability to perceive them as an integral unit (thereby reducing the perception of workload) and to reduce the need to switch information obtained from one sensory modality to subsequent activities displayed in another.

REFERENCES

- Fitts, P. M. & Peterson, J. R. (1964) Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 67, 103-112.
- Gopher, D. & Braune, R. (1984) On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, 26, 519-532.
- Gopher, D., Brickner, M. & Navon, D. (1982) Different difficulty manipulations interact differently with task emphasis: Evidence from multiple resources. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 146-157.

- Hart, S. G., Sellers, J. J. & Guthart, G. (1984) The impact of response selection and response execution difficulty on the subjective experience of workload. In *Proceedings of the 28th Annual Meeting of the Human Factors Society*. Santa Monica, CA: Human Factors Society, 732-736.
- Hartzell, E. J., Gopher, D., Hart, S. G., Dunbar, S. & Lee E. (1983) The Fittsberg Law: The joint impact of memory load and movement difficulty. Paper presented at the 27th Annual Meeting of the Human Factors Society. Seattle, WA.
- Kantowitz, B. H. & Knight, J. L. (1978) Testing tapping, time-sharing: Attention demands of movement amplitude and target width. In G. E. Stelmach (Ed.), *Information Processing in Motor Control and Learning*. New York: Academic Press, 205-227.
- Navon, D. & Gopher, D. (1979) On the economy of the human processing system. *Psychological Review*, 86, 214-255.
- Mosier, K. & Hart, S. G. (1985) Levels of Information Processing in a Fitts Law Task (LIP-Fitts), This volume, in press.
- National Research Council (1982) *Automation in Combat Aircraft*. Washington, D. C. National Academy Press.
- Pachella, R. (1974) The interpretation of reaction time in information processing in research. In B. H. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition*. : Lawrence Erlbaum Associates, Hillsdale NJ. 41-82
- Remington, R. (1977) Processing of phonemes in speech: A speed-accuracy study. *Journal of the Acoustical Society of America*, 62, 1279-1290.
- Staveland, L., Hart, S. G. & Yeh, Y.- Y. (1985) Memory and Subjective Workload Assessment, This volume, in press.
- Sternberg, S. (1969) The discovery of processing stages: Extension of Donder's method. *Acta Psychologica*, 30, 276-315.
- Tsang, P. S. & Wickens, C. D. (1984) The effects of task structures on time-sharing efficiency and resource allocation optimality. *Proceedings of the 20th Annual Conference on Manual Control NASA CP-2341*. Washington, D. C.:National Aeronautics and Space Administration, 305-318.
- Vidulich, M. A. & Tsang, P. S. (1985) Techniques of subjective workload assessment: A comparison of two methodologies. *Proceedings of the Third Biannual Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, in press.
- Vidulich, M. A. & Wickens, C. D. (1985) Stimulus-central processing- response compatibility: Guidelines for the optimal use of speech technology. *Behavior Research Methods, Instruments, and Computers*, in press.
- Wickens, C. D. & Kessel, C. (1979) The effects of participatory mode and task workload on the detection of dynamic system failures. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9, 1080-1082.
- Wickens, C. D., Sandry, S. L., & Vidulich, M. A. (1983) Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227-248.

Wickens, C. D. & Yeh, Y.- Y. (1982) The dissociation of subjective workload and performance. *Proceedings of the International Conference on Cybernetics and Society*. New York: Institute of Electrical and Electronics Engineers, 584-587.

Wickens, C. D. & Yeh, Y.- Y. (1983) The dissociation between subjective workload and performance - A multiple resources approach. *Proceedings of the Human Factors Society 27th annual Meeting*. Santa Monica, CA: Human Factors Society, 244-248.